IDENTIFYING PERTINENT VARIABLES FOR NON-RESPONSE FOLLOW-UP SURVEYS: LESSONS LEARNED FROM FOUR CASES IN SWITZERLAND.

Caroline Vandenplas[1], Dominique Joye[2], Michèle Ernst Staehli[3], and Alexandre Pollien[3]

Septembre 2015

[1]Corresponding author:

Centre for Sociological Research

Katholieke Universiteit Leuven

Parkstraat 45 - box 3601

BE-3000 Leuven Belgium

caroline.vandenplas@soc.kuleuven.be

[2]Institute of Social and Political Sciences

University of Lausanne

Géopolis-5548

CH-1015 Lausanne Switzerland

Abstract

All social surveys suffer from different types of errors, of which one of the most studied is non-response bias. Non-response bias is a systematic error that occurs because individuals differ in their accessibility and propensity to participate in a survey according to their own characteristics as well as those from the survey itself. The extent of the problem heavily depends on the correlation between response mechanisms and key survey variables. However, non-response bias is difficult to measure or to correct for due to the lack of relevant data about the whole target population or sample. In this paper, non-response follow-up surveys are considered as a possible source of information about non-respondents. Non-response follow-ups, however, suffer from two methodological issues: they themselves operate through a response mechanism that can cause potential non-response bias, and they pose a problem of comparability of measure, mostly because the survey design differs between main survey and non-response follow-up. In order to detect possible bias, the survey variables included in non-response surveys have to be related to the mechanism of participation, but not be sensitive to measurement effects due to the different designs. Based on accumulated experience of four similar non-response follow-ups, we studied the survey variables that fulfill these conditions. We differentiated socio-demographic variables that are measurement-invariant but have a lower correlation with non-response and variables that measure attitudes, such as trust, social participation, or integration in the public sphere, which are more sensitive to measurement effects but potentially more appropriate to account for the non-response mechanism. Our results show that education level, work status, and living alone, as well as political interest, satisfaction with democracy, and trust in institutions are pertinent variables to include in non-response follow-ups of general social surveys.

Introduction

Response rates in social surveys have been decreasing in recent decades (Brick & Williams, 2013; De Leeuw & De Heer, 2002). This decline in response propensities can at least partially be explained by a non-favourable survey climate (Lorenc, Loosveldt, Mulry, & Wrighte, 2013) that can be induced by the large amount of surveys conducted in Western-European countries, including marketing research. Even though it has been thoroughly discussed that response rates are not directly linked to non-response bias (Groves & Peytcheva, 2008) and that blindly trying to increase the response rate can have an unexpected effect on bias (Beullens & Loosveldt, 2012), lower response rates increase the risk of non-response bias. Costly efforts are implemented during the data collection process of social surveys to counter-balance this effect, e.g., increased contact attempts, refusal conversion, monetary incentives (Stoop, Billiet, Koch, & Fitzgerald, 2010), and even targeted (Schouten, Schlomo, & Skinner, 2011) or responsive design (Groves & Heeringa, 2006). A vast literature also exists about post-hoc non-response adjustment through weighting and its limitations (e.g., Kalton & Flores-Cervantes, 2003; Särndal & Lundström, 2010; Schouten, 2007; Skinner & D'Arrigo, 2011). The challenge facing survey methodologists nowadays is the necessity to understand the mechanism of (non-)response and the causes of non-response bias in depth (Brick, 2014). Key to this understanding is to collect information about non-respondents that is both reliable and pertinent. So far, two sources of information on non-respondents have been proposed, using paradata (Kreuter & Olson, 2010), or auxiliary data (e.g., Sinibaldi, Trappmann, & Kreuter, 2014), and conducting a non-response follow-up survey (NRS) (Hansen & Hurwitz, 1946).

Paradata (Kreuter, 2010) exist in many forms, such as observable data (Lynn & Nicolaas, 2010; Matsuo & Billiet, 2009; West, 2013) or data about the survey process itself (Beaumont, 2005; Pollien & Joye 2014; Wagner & Hubbard, 2014). They have the advantage of existing for all respondents and non-respondents, but the information available is rather restricted and the relation with survey items can be weak, especially in general social surveys

NRS are surveys dedicated to non-respondents, usually consisting of a short questionnaire and a change of mode, to try to convince a large number of non-respondents to participate. Hansen and Hurwitz (1946) first proposed the idea of "call-back" surveys — surveys amongst refusals using the entire questionnaire. Later, a variant known as the "basic question procedure" was introduced where the non-respondents (refusals and/or non-contacts) are asked to complete or (re-) approached shortly after the end of the main data collection with a shorter questionnaire containing only one or a few questions (Bethlehem & Kersten, 1985). An intermediate type of NRS is the PEDASKI (Pre-emptive Doorstep Administration of Key Survey Items) method that was applied to the British Crime Survey (Lynn, 2003). In this case, the basic idea is that the survey interviewer, having made contact with a sample member, should ask a small number of survey items as soon as it becomes apparent that no interview is going to be achieved.

The advantage of NRS is that the variables included in the questionnaire can be chosen to be in relation with response propensities and variables of interest. However, it loses the benefit of having information about the whole sample, as there may still be a group of 'repeated[1]' non-respondents that do not participate in either the main survey or the NRS. The extent of this issue depends on how far these 'repeated' non-respondents differ from the non-respondents taking part in the NRS. This difference can be evaluated with paradata or frame data (Roberts, Vandenplas, & Ernst Staehli, 2014). To achieve as high as possible response rates for non-response follow-ups, such surveys have to be short while still gathering useful information. It is therefore important to identify the most pertinent items to include in the questionnaire. Moreover, the survey designs of the main and the non-response follow-up often differ — e.g., different time-frame, mode, questionnaire — leading to different measurement effects. Therefore, variables included in a NRS should have the following properties:

- Being **measurement invariant** between the main and non-response survey. If the design of the non-response follow-up calls for a change of mode, variables included in the NRS should be relatively mode insensitive to avoid the detected differences

response follow-up is conducted after the main survey (see further for more details about the specific implementation), a shift due to the time elapsed should be avoided. Moreover, the data collection of the main survey being spread over more than one month, such time sensitive variables could suffer from less precision due to time variation (Saris & Galhoffer, 2007, Chapter 10). The answer to some items may also be conditioned in the main survey by the context or the module from which they originate. Non-response follow-up variables should therefore not be subject to such conditioning.

- Being **pertinent for the detection of non-response bias**. This means being related to (as many as possible) variables of interest (Groves, 2006; Kalton & Flores-Cervantes, 2003; Kalton & Maligalig, 1991; Little, 1986; Little & Vartivarian, 2003, 2005) and covering all the dimensions of non-response. The non-response mechanism is underlined by several processes; at a basic level, we can distinguish refusal and non-contact. Some variables are more related to the refusal process, while others more to the non-contact. Ideally, the set of variables used in a non-response survey must cover both aspects.

In this paper, a collection of surveys and their non-response follow-ups conducted in the last 10 years in Switzerland will be analysed. Variables included in the non-response follow-ups will be tested for measurement-invariance across the main survey and non-response follow-up and for their pertinence in detecting non-response bias.

Data

In this section, the four considered cross-sectional international surveys and their non-response follow-ups conducted in Switzerland are described:

- The European Social Survey: round 3, 2006 and round 5, 2010 (ESS)
- The European Value Study, 2008 (EVS)
- Measurement and Observation of Social Attitudes in Switzerland 2011, including ISSP

Figure 1
General description of the surveys

ESS 2006	EVS 2008	ESS 2010	MOSAiCH 2011
		Repeated	
Every 2 years	Every 9 years	Every 2 years	Every 2 years
		Topic	
political and economic sciences, sociology, and social psychology	basic values, attitudes, and preferences of the European population	political and economic sciences, sociology, and social psychology	two editions of ISSP plus questions concerning the relation between Switzerland and Europe and questions concerning Swiss political institutions.
		Target population	
15 or more years old country residents	18 or more years old country residents	15 or more years old country residents	18 or more years old country residents
		Sampling frame	
telephone directory, household	list of dwellings provided by the Swiss Post	Swiss Federal Statistical Office (SFSO) register, individual	Swiss Federal Statistical Office (SFSO) register, individual
		Sample size	
3,713 households, one person in responding household	2,970 dwellings, one person in responding household	2,850 individuals	2,409 individuals
		Contact procedure	
up to 5 face-to-face contact attempts	up to 5 face-to-face contact attempts	up to 5 face-to-face contact attempts	up to 5 face-to-face contact attempts
		Phone contacts	
Yes, at the end of the fieldwork	No	Yes, at the end of the fieldwork	Yes, at the end of the fieldwork
		Incentives	
30 Sfrs (24 Euros) conditional (cash, voucher, or donation)	30 Sfrs (24 Euros) conditional (cash, voucher, or donation)	30 Sfrs (24 Euros), 4/5 of the sample conditional (cash, voucher, or donation), 1/5 prepaid cash	30 Sfrs (24 Euros) conditional (cash, voucher, or donation)
		Mode and length of interviews	
CAPI, 60 to 70 min	CAPI, 60 to 70 min	CAPI, 60 to 70 min	CAPI, 60 to 70 min
		Duration of fieldwork	
32 weeks	22 weeks	26 weeks	36 weeks
		Refusal conversion	
yes	yes	yes	yes

We can observe that, on the one hand, the topic, sampling frame, sampling design, and

A description of the MOSAiCH variables can be found on the FORS website[4].

Table 1 gives a description in terms of fieldwork outcomes for each study: the number of completed interviews, ineligibles, refusals, and non-contacts, as well as other categories of non-respondents and response rates.

TABLE 1

Number of sampled units, completed interviews, ineligibles, refusals, and non-contacts in the main survey

	ESS2006	EVS2008	ESS2010	MOSAiCH2011
Sample units	3,710	2,970	2,850	2,409
Ineligibles	257	113	37	129
Completed interviews	1,804	1,271	1,506	1,212
Refusals[b]	1,446	1,033	871	697
Non-contacts	150	553	200	138
Other non-respondents[c]	53	65	236	233
Response rates (%) [d]	52.2	44.5	53.5	53.2

[a]Ineligibles: deceased, moved to another country, address not traceable, address not residential, other ineligibles.
[b]Refusals: Refusals by respondent, by proxy or household refusal, partial interview, broken interview.
[c]Other non-respondents: not available/away, contact but no interview, mentally or physically unable, language barrier, moved to unknown address, moved still in country.
[d]Response rates are calculated based on the number of sampled units minus the number of ineligibles. They are equivalent to AAPOR Response Rate 1.

We can observe in Table 1 that the EVS had the lowest response rate, probably due to a reduction of the data collection budget. The number of ineligibles was very low in the ESS 2010, probably due to the population register frame, while MOSAiCH, which benefited from the same frame, had many ineligibles. However, this could be due to the timing of the sampling procedures, as the frame is only updated every three months. In ESS 2006, the non-contact (and other non-respondents) rate was exceptionally low. This is to be explained by the use of the telephone register as a sampling frame. The telephone register probably failed to include the entire population, replacing the non-contact problem by non-coverage. These results underline the relation between the sampling frame, the non-response rate, and the responding sample composition in a total survey error perspective.

Non-response follow-up surveys

The non-response follow-ups were designed as short self-administered paper

sample unit. The latter is quite complex, as the main survey includes different re-issuing procedures for refusals, non-contacts, and other non-respondents. The time frame between the initial contact for the main data collection and receiving the non-response follow-up questionnaire could hence vary from two to six months depending on the time of the initial contact.

For each study, the target was all non-respondents as well as a control group of 200 (EVS 2008 and MOSAiCH 2011) or 300 respondents (ESS 2006 and 2010) to the main survey. See Figure 2 for an overview of the different groups of respondents and non-respondents to the main survey and NRS. The contact letters were targeted to the type of unit. The participation or non-response to the main survey as well as the reason for non-participation was acknowledged in the contact letter (respecting the refusal, regretting the non-contact, thanking again for participation, etc.). It was also explained that the responses to this short survey were needed in order to conclude the project and ensure the quality of the data of the main survey. The questionnaires were sent together with a 10 Sfrs (8 Euros) unconditional incentive. For all the non-respondents who did not return the NRS questionnaire after approximately three weeks, a telephone reminder was organised during which the questionnaire could be answered through a CATI (computer-assisted telephone) interview. If no phone number was available, a paper reminder was sent by mail.

The French version of the non-response questionnaires can be found in Appendix B. Some variables have been recoded; see Appendix C for more details.

Figure 2
Different groups of respondents and non-respondents to the main and non-response follow-up surveys.

[Figure 2 to be inserted here]

Table 2 summarises the number of persons who received a NRS questionnaire and returned it or answered by phone.

Invalid	25	9	1	37
Refusals	714	486	464	371
Non-contacts	15	256	110	60
Other NR	5	11	8	6
Response rates to NRS (%)	56.3	49.7	63.3	62.4
Cumulative RR without ineligibles (%)	73.5	70.8	74.2	72.3

[a]No non-response survey questionnaires were sent to sampled individuals who were deceased, moved out of the country, or had explicitly expressed that he/she did not want to be contacted anymore (per letter or phone). See Appendix A for a more in-depth study of why some non-respondents did not receive the NRS questionnaire

For each considered survey, about half of the non-respondents participated in the non-response survey, which is a considerable gain, at least when considering the response rate increase. This relatively high response rate for a non-response follow-up can be attributed to the change of mode, the shortness of the questionnaire, and the incentive sent out with the NRS (10 Sfrs/8 Euros).

The choice of variables for the non response follow ups.

The non-response follow-ups contain questions extracted from the corresponding main surveys including attitudinal and socio-demographic variables. The selected variables were suspected and partly confirmed to be related to response propensity (e.g., Matsuo, Billiet, Loosveldt, Berglund, & Kleven, 2010). The variables included in the non-response follow-ups were chosen to cover the different aspects of the response mechanism, including accessibility and cooperation.

In face-to-face surveys, variables such as time spent at home, practical accessibility of the dwelling, and personal accessibility of the persons in their dwelling are believed to be linked to the propensity to be contacted. Information about geography, having a partner, working time, and social activities can operate as proxies.

Moreover, the propensity of a contacted person to cooperate depends on his/her social disposition to be part of the social group of respondents. This disposition is related to the relationship between respondents and the public spheres, which embodies, among other things, the ability and legitimacy to form and express an opinion. Concretely, variables measuring political interest, attitudes toward foreigners, or trust in others should be related to

All analyses were performed in R.2.15[6]. The sampling weights were not taken into account, because the aim was not to produce population estimates but to compare groups. Moreover, the sampling weights for all surveys were exactly or very close to 1 due to the balanced sampling designs.

Measurement-invariance of the items measured in the non-response surveys

A variable is measurement-invariant if its measurement in the paper/CATI questionnaire is comparable to its measurement in the main questionnaire.

Differences in measurement effects can be systematic or random. Both types, systematic and random, are to be avoided, although systematic differences are worse as they can induce wrong conclusions; detected differences between respondents and non-respondents to the main surveys are actually due to differences in measurement errors rather than to non-response bias in the main survey. To detect systematic differences in measurement between the surveys, differences in distribution of answers given to the main survey and to the non-response follow-up by the control group were tested using the Wilcoxon test for the ordinal variables and the Pearson Chi square test for the dichotomous variables. An important assumption was made when using the control group — namely, that the answering behaviour in the NRS is the same for respondents (control group) and non-respondents to the main surveys. We also assumed no conditioning effect from having participated in the main survey for the control group. This last assumption seems reasonable, given that at least two months separate the face-to-face interview and the filling in of the NRS questionnaire.

To test in how far the assumption that the control group and non-respondents to the main survey have the same answering behaviour in the non-response follow-up holds, we tested differences in measurement through another method. We weighted the control group to the full group of respondents to the non-response follow-up using a propensity score weighting based on all the non-response follow-up variables. We then compared the answers given to the main survey by the control group, weighted or not, to assess bias. The weighted

Random differences do not cause bias and are therefore less problematic. However, variables for which random differences in measurement effects would occur between the main and non-response survey would not be well suited for the detection of non-response bias in the main survey. The random difference would increase the standard error of the variable, lowering the significance of the differences between respondents and non-respondents to the main survey. The aim was thus to exclude such variables from non-response follow-ups. To measure the reliability of the non-response follow-up variables, we calculated the correlation of the answers given to the different surveys by the same respondent divided by the reliability of the main question itself. Lacking the latest measure for most variables, we decided to stick to the correlations between the main and NRS answer, mimicking the idea of test-retest design to measure the 'reliability' applied in the ESS (Saris, Oberski, Revilla, Zavala, Lilleoja, Gallhofer, & Gruner, 2011). To give some perspective to this correlation, we compared it to the reliability of the item with the so-called retest-item that has been re-included at the end of the questionnaire whenever possible (ESS variables). Experiments testing the reliability of the survey variables have been included in each ESS questionnaire since the first round (Saris et al., 2011).

The response rates of the control groups varied between 80% and 90% (see Table 2). These very high response rates are not surprising, as the studied groups are groups of people who already agreed to participate in the main survey, which requires a bigger time investment for the participant than filling in the shorter non-response questionnaire. However, we have to keep in mind that the control groups are quite small (159 to 267).

Each NRS contains two types of variables: general socio-demographic and variables thought to be correlated to non-response both generally and survey-specific. Table 3 shows the mean differences (significance level tested with Pearson's Chi square test) and the correlation (Phi coefficients) of the socio-demographic variables across surveys. All variables are dichotomous.

	ESS 2006			EVS 2008			ESS 2010			MOSAiCH 2011		
	n	ρ	diff	n	ρ	diff	n	ρ	diff	n	ρ	diff
Gender	126	1	0	154	1	0	253	1	0	175	1	0
Age (3 categ.)	126	1	-0.01	153	1	-0.02	254	1	0.02	174	1	0
High education	237	0.82	0.03	156	0.79	0.41	255	0.79	0.18	175	0.85	0.12
Employed	238	0.80	0.01	153	0.54	-0.01	257	0.79	-0.03	175	0.80	-0.02
Full time empl.										157	0.58	0.05
Living alone	228	0.86	-0.04	131	0.86	-0.01	245	0.84	0.00	172	0.88	-0.01
Having a partner										132	0.48	0.00
Having children										169	0.96	0.01

The socio-demographic variables have a high correlation (>0.45) between the main survey and NRS, with the lowest in MOSAiCH being ‘having a partner’ and ‘working full time’, which can change over time. The tests of equal median or equal proportion do not reject the hypothesis that the distributions are the same. In ESS 2010, however, the distribution between the education variable measured in the main and NRS differ with a low significance (<0.1). An explanation for such a low correlation is that the categories had changed from the main to the non-response follow-up, as all the categories offered in the face-to-face interview to fit the international coding scheme (Schneider & Kogan, 2008) could not be included in the paper version. In general, as expected, the socio-demographic variables are not or are only slightly affected by measurement effect.

In Table 4, mean differences and correlations of the answer given to the questions repeated in the non-response follow-up across surveys are displayed for attitudinal variables that are likely to have a direct relationship with participation in surveys. Almost all variables are ordinal. The dichotomous variables are ‘landline phone’, ‘mobile phone’, ‘registered number’, and ‘group membership’. Spearman correlation coefficients were calculated in the case of ordinal variables and Phi coefficients for dichotomous variables. The significance

More than half of the considered variables cannot be considered as measurement-invariant. Variables that do not have systematic differences in measurement are ‘political interest’(except in MOSAiCH), ‘satisfied with democracy’ (only ESS 2010), ‘trust in legal system’, ‘duty to inform authorities’, ‘science can solve environmental problems’, ‘science has positive effects’, ‘trust in health system’, ‘trust in educational system’, ‘importance of work, family, politics and religion’, ‘social activities’(only ESS 2010), ‘worry about environment’, ‘having influence on events’, ‘tired after work’, ‘believe surveys are useful’ (only in ESS 2010), ‘landline, mobile’, and ‘registered number’. All other variables displayed a significant shift between the answer given during the main survey and the non-response follow-up questionnaire. Most of these shifts can be explained by a social desirability effect due to the presence of an interviewer in the main survey, e.g., immigration or trust variables. The method using the weighted control group detected fewer measurement differences in general. However, it detected some significant differences for ‘political interest’ (p-value<0.05) and ‘believe surveys are useful’ (p<0.001) in ESS 2010 and ‘importance of work’ in EVS (p<0.05). It also never detected any significant non-response bias between the un-weighted answers and weighted answers of the control group, which seems to suggest that weighting did not account completely for the selection effects.

From the variables not having systematic differences, the correlation for the items ‘political interest’ and ‘satisfied with democracy’ is high[9] for all surveys, except in MOSAiCH. This exception could be due to an inversion of the response scale between the main survey and the NRS or to external political events in the time elapsed between the main survey and the NRS. The reliability (Saris & Gallhofer, 2007, p 190–191) of ‘satisfied with democracy’ measured in Switzerland in different rounds of the ESS and in different languages (German, French, and Italian) varied between 0.80 and 0.86. ‘Political interest’ was never tested as a measure in Switzerland; a test in the UK in ESS round 5, however, showed a reliability of 0.70.

TABLE 4

Number of cases considered, correlations, and mean differences between the answers given to the main survey and the NRS by the control group. Variables linked to participation.

	ESS 2006			EVS 2008			ESS 2010			MOSAiCH 2011		
	n	diff	ρ	n	diff	ρ	n	diff	ρ	n	diff	ρ
Political interest	242	0.01	0.61	156	-0.20	0.60	257	-0.03	0.67	174	-0.14***	0.68
Satisfied with democracy	122	0.38	0.44				254	-0.16	0.50	170	0.13***	0.37
Immigration	125	-0.04	0.38	154	-0.23*	0.63	256	-0.29**	0.56	168	0.57***	0.55
Trust in politicians	125	-0.32	0.47									
Duty to inform authorities				150	-0.02	0.27						
Trust in legal system							254	0.13	0.55			
Accept police decision							250	-0.01	0.37			
Science can solve environmental problems							251	-0.12	0.30			
Science has positive effects										171	0.07	0.11
Trust in health system										171	0.06	0.40
Trust in educ. system										163	0.07	0.37
Trust in other people	125	-0.28	0.50	153	-0.40*	0.35				173	-0.16	0.47
Feel safe after dark	237	-0.09	0.54				255	0.26***	0.39			
Social activity	236	0.15***	0.45				252	0.00	0.46			
Meeting socially frequently							254	-0.30**	0.44			
Watch TV	126	0.26*	0.75									

Voluntary work	125	-0.58***	0.64										
Importance of work				152	0.00	0.46							
Importance of family				154	0.01	0.33							
Importance of friends				155	0.10	0.38							
Importance of leisure				153	0.18**	0.42							
Importance of politics				154	-0.05	0.58							
Importance of religion				151	0.06	0.75							
Stick to own affairs				154	0.32	0.36							
Group membership				145	0.30***	0.24							
Worry about environment										171	0.09	0.44	
Help in daily tasks										169	0.60**	0.36	
Emotional support										172	-0.16	0.40	
Happy							258	-0.32**	0.40	175	0.36***	0.33	
Health							259	0.14***	0.65	175	0.43***	0.38	
Worry about work							131	0.30**	0.37				
Tired after work							131	-0.016	0.21				
Having influence on events										175	-0.04	0.25	
Easily solve problems										166	0.20	0.16	
Believe surveys useful							246	-0.01	0.36	161	0.16	0.17	
Landline phone							256	-0.02	0.55	175	0.02	0.75	
Mobile phone							252	0.01	0.59	171	0.05	0.13	

Registered number (landline or mobile)	211	-0.02	0.67	148	-0.04	0.62

* p<0.05, ** p<0.01, and p<0.001.

In ESS 2010, the item ‘trust in legal system’ is robust against the change of survey design, concurring with the ESS reliability measure (more than 0.8).

The item ‘social activity’ also seems to not suffer from random measurement difference, similarly to the EVS 2008 items ‘importance of work’, ‘importance of politics’, and ‘importance of religion’. Finally, answers about whether one has a landline phone, mobile phone, or registered telephone number seem measurement-invariant across the survey designs except for the mobile phone in MOSAiCH. These strong correlations probably reflect the factual character of these questions, even if they might represent a proxy of some lifestyle and attitude dimensions.

Predictive power of the non-response survey items

We now have identified the measurement-invariant variables. For these variables to be useful, they also need to predict non-response bias, i.e., if respondents and non-respondents participating in the non-response survey differ.

For each survey in the following tables (5 to 8), we used a Pearson Chi square test to define whether the items discriminate for participation, which is a key feature for items included in a non-response survey. The distribution of the measurement-invariant variables over the respondents to the main survey and the non-respondents to the main survey who participated in the non-response follow-up are displayed, as well as the difference also often called contrast. We decided to not stick to mean or median differences because differences between respondents and non-respondents were sometimes more subtle. No difference in mean does not mean there is no difference in distribution. For the main purpose of increasing the readability, some variable categories were put together, provided that they displayed the same tendency between respondents and non-respondents.

TABLE 5

Percentage estimate from ESS 2006 for respondents and non-respondents who completed the non-

Living alone (n=1,803/228)	30.9	26.2	-4.7
Political interest (n=1,801/242)			
1– not interested at all	15.7	11.4	-4.3*
2 – a little interested	42.1	43.5	1.4*
3 – quite interested	31.6	34.8	3.2*
4 – very interested	10.7	10.3	-0.4*

* p<0.05, ** p<0.01, and ***p<0.001.

[a] Primary, Secondary, Apprenticeship

In the ESS 2006, the socio-demographic variables do not show any significant differences between respondents and non-respondents (participating to the non-response survey). 'Political interest' is significantly discriminating, non-respondents reporting less often extreme values -- being not interested at all or very interested in politics -- and more often being a little or quite interested in politics than respondents to the main survey. Note that some variables were only included in half of the non-response questionnaire due to an experiment on the impact of the length of the questionnaire on response rate for the NRS. This could be the reason why no significant differences were found between respondents and non-respondents for some of the variables.

Table 6 displays the frequency per category of the variables that have been found to be reliable by the previous analysis for EVS 2008.

Table 6

Percentage estimate from EVS 2008 for respondents and non-respondents who completed the non-response questionnaire for the items that were shown to be stable in the previous section.

	Respondents	Non-respondents	Contrast
Gender (Male)(n=1,271/154)	46.1	49.3	3.2
Age (n=1,271/153)			
<30	14.8	15.7	0.9
>30, <65	62.8	62.0	-0.8
>65	22.4	22.3	-0.1
Education[a] (n=1,260/156)	64.8	49.3	-15.5***
Employed (n=1,261/153)	63.6	70.5	6.9***
Living alone (n=1,005/131)	18.4	29.7	11.3***
Political interest (n=1,266/156)			
1– not interested at all	14.4	11.9	-2.5***
2 – a little interested	33.8	46.8	13.0***
3 – quite interested	34.2	31.6	-3.2***
4 – very interested	17.6	9.7	-7.9

Importance of religion (n=1,253/151)			
very important	17.1	15.0	-2.1
rather important	26.1	24.0	-2.1
not very important	34.7	35.9	1.2
not important at all	22.1	25.1	3.0

* p<0.05, ** p<0.01, and ***p<0.001.

[a] Primary, Secondary, Apprenticeship

Respondents to EVS 2008 were significantly more likely to be better educated but less likely to be employed and to live alone than non-respondents. Respondents were also more likely to be quite or very interested in politics and less likely to be a little interested than non-respondents. But a higher proportion of respondents reported not being interested at all in politics. Other variables that seem to differentiate respondents and non-respondents are 'importance of work' and 'importance of politics'; here again there is no distinction of important/not important for respondents versus non-respondents. Respondents more often reported work as being very important, not very important or not important at all than non-respondents. At the same time, respondents reported less often work as being rather important than non-respondents. Politics was more often reported as very important or not important at all (extreme values) by respondents than non-respondents, but less often as being rather or not very important.

TABLE 7

Percentage estimate from ESS 2010 for respondents and non-respondents who completed the non-response questionnaire for the items that were shown to be stable in the previous section.

	Respondents	Non-respondents	Contrast
Gender (Male)(n=1,506/253)	51.3	45.2	-6.1*
Age (n=1,502/254)			
<30	20.8	18.5	-2.3
>30, <65	60.4	62.0	1.6
>65	18.8	19.5	0.7
Education[a] (n=1,505/255)	40.3	45.9	5.6*
Employed (n=1,481/257)	57.7	57.1	-0.6
Living alone (n=1,506/245)	17.9	20.0	2.1
Political interest (n=1,501/257)			
1– not interested at all	16.1	10.5	-5.6**
2 – a little interested	42,7	47.7	5.0**
3 – quite interested	30.4	32.7	2.3**
4 – very interested	10.7	9.1	-1.6**
Satisfied with democracy, >7 (n=1,473/254)	69.7	57.4	-12.3***
Trust in legal system (n=1,469/254)			

Landline phone (n=1,506/256)	90.6	83.7	-6.9***
Mobile phone (n=1,501/252)	87.7	88.7	1.0
Registered number (n=1,349/211)	94.9	81.9	-13.0***

*p<0.05, ** p<0.01, and ***p<0.001.
[a] Primary, Secondary, Apprenticeship

Non-respondents to the ESS 2010 were less likely to be male and satisfied with democracy and more likely to be low educated. They were also less likely to report no interest at all or very high interest in politics (extreme values) but more likely to report that they were a little or quite interested in politics than respondents. Non-respondents were more likely to report meeting socially a lot less, about the same and more often than most than respondents to the main survey. Non-respondents were also less likely to report meeting socially less or a lot more often than most than respondents. The difference in the answer pattern for the variable 'social activity' is hence also showing a non-linear pattern. Finally, they were less likely to have a landline phone or a registered number.

TABLE 8

Percentage estimate from MOSAiCH 2011 for respondents and non-respondents who completed the non-respondents questionnaire for the items that were shown to be stable in the previous section.

	Respondents	Non-respondents	Contrast
Gender (Male)(n==1,212/175)	50.8	43.8	-7.0***
Age (n=1,212/126)			
<30	17.5	15.8	-1.7
>30, <65	63.1	66.4	3.3
>65	19.4	17.7	-1.7
Education[a] (n=1,210/175)	60.0	47.2	-12.8***
Employed (n=1,210/175)	61.3	66.3	5.0
Full time employment (n=1,088/157)	73.6	83.1	9.5***
Living alone (n=1,212/172)	19.1	18.0	-1.1
Having a partner (n=866/132)	93.6	66.2	-27.4***
Having children (n=1,212/129)	29.5	33.2	3.7
Landline phone (n=1,202/175)	88.3	82.7	-5.6**
Registered number (n=1,069/148)	90.8	74.7	-16.1***

* p<0.05, ** p<0.01, and ***p<0.001.
[a]Primary, Secondary, Apprentice

In MOSAICH 2011, non-respondents were less likely to be male, to be low educated, and to have partner, and more likely to be employed and have a full-time job. They were also less likely to have a landline phone or a registered phone number.

invariant and able to predict non-response bias to be deemed fit for inclusion in non-response follow-ups.

The items directly linked to non-response were assumed to measure or be proxies for lifestyle, trust, social participation, or integration in the public sphere. The analyses show that the socio-demographic items do not suffer from or are only slightly affected by the changes of mode, protocol, and the time elapsed between the main and the non-response surveys. Gender was found to be discriminating for ESS 2010 and MOSAiCH 2011, with non-respondents more likely to be women. Non-respondents were also less likely to be low educated in EVS and MOSAiCH and more likely to be better educated in ESS 2010. They were also more likely to be employed (EVS and MOSAiCH), work full time (MOSAiCH), and live alone (EVS), and less likely to have a partner (MOSAiCH). Hence, gender, education level, employment status, and household situation seem to be variables that should be included in non-response follow-ups.

Considering attitudinal variables, the items 'political interest', 'satisfied with democracy', 'social activity', 'trust in legal system', or 'importance of politics and work' were found to be measurement-invariant, at least in most cases when the items were present in different non-response follow-ups, and pertinent to detect non-response bias.

The use of the control group to test measurement-invariance implies the assumption that the respondents and non-respondents to the main survey had the same answering behaviour in the non-response survey. In other words, we assumed that differences in measurement errors between the main and non-response surveys are independent of the response mechanism, which is not trivial. This assumption was tested using another method to detect measurement difference, namely comparing the weighted answers to the main survey of the control group with the non-respondents' answers to the non-response follow-up. From the above-listed measurement-invariant variables, only 'importance of work' was detected as having a systematic measurement difference by this second method.

extreme interest in politics. Other variables displayed similar patterns, including ‘social activity’, ‘importance of work’, and ‘importance of politics’. The findings of these patterns of differences in distribution between respondents and non-respondents fits with the idea that respondents are people that have an opinion, either positive or negative. The respondents are commonly integrated into the public sphere and not afraid of taking a position, therefore more likely to give more extreme values as answers than non-respondents. The variables about trust (‘trust in people’, ‘trust in legal system’) showed that the declaration of trust in the course of a survey appears as a key measure of propensity to participate. Non-respondents trust the legal system less (ESS 2010). In this heteroclite but general distrust, we can again read a kind of relation to the social environment. Non-response surveys seem to show that respondents are stable persons who are integrated into the public sphere.

Finally, non-respondents were also less likely to have a landline phone or a registered phone number (ESS 2010 and MOSAiCH). Although this cannot be interpreted as a social mechanism, these variables seem to be a proxy for the social inclusion discussed above.

The results show indicators that are discriminatory for all examined surveys; others are not always discriminating. This instability suggests that the participation process is complex and depends on and interacts with several independent dimensions related to the strategies put on refusal conversions, finding non-contacts, and the topic of the survey.

Reflecting on the experience with non-response follow-ups, some improvement could be brought to the design of these surveys. To minimise measurement effects, the question design in the non-response follow-up should mimic as much as possible the one in the main survey. It is important to avoid scale inversions or changes in wording, despite the mode change. The time elapsed after the main survey should also be minimised to prevent external factors such as political, public, personal, or even seasonal events from influencing the answer given to variables repeated in the non-response follow-ups. However, this could be a problem for the use of the control group as it could increase conditioning effects. Another possibility to ameliorate the design is to consider including more than one variable on the same construct to

The differences between respondents and non-respondents found in the results could be specific to non-respondents participating in the non-response follow-up. The group of repeated non-respondents could display other characteristics than the ones found amongst non-respondent that participated to the non-response survey. Indeed, using non-response follow-ups for detection and correction of non-response bias is based on two assumptions. The first one, the measurement-invariance of the variables included in the NRS, has been extensively discussed and controlled for in this paper. The second one has only been briefly mentioned. The non-response follow-ups also suffer from non-response bias; a group of 'repeated' non-respondents do not participate in either the main survey or the NRS. Using the non-response survey implies the assumption that the non-respondents participating in the NRS are representative of all non-respondents, including 'repeated' non-respondents. In ESS 2010, a population register was used as a sampling frame. This allowed Roberts et al. (2014) to study the differences in some socio-demographic variables of these 'repeated' non-respondents. The authors found that, amongst respondents to the main survey and the NRS, representing approximatively 75% of the sample, Swiss citizens as well as people having a registered telephone number and living in rural areas were overrepresented as it is the case if we limit ourselves to the group of respondents to the main survey.

Despite the methodological challenges encountered by non-response follow-ups, compliance to the measurement-invariance and non-respondents representativity, they are good tools to comprehend the non-response mechanism. Identifying good participation indicators is a key step in detecting and treating non-response and non-response bias. The non-response surveys showed which indicators represented strong and hence more stable attitudes and were differentiating respondents and non-respondents to the main survey at the same time. One of the main issues in dealing with non-response is the lack of information about non-respondents. The non-response surveys, in the Swiss case, give relevant information about approximately half of the non-respondent sample, which is not negligible.

REFERENCES

Beullens, K., & Loosveldt, G. (2012). Should high response rate really be a primary objective? *Survey Practice*, *5*(3), Retrieved June 21, 2014 from http://surveypractice.org/index.php/SurveyPractice/article/view/21/html.

Brick, J. M. (2014). Unit Nnnresponse and weighting adjustments: a critical review. *Journal of Official Statistics*, *29*(3), 329–353.

Brick, J. M., & Williams, D. (2013). Explaining rising nonresponse rates in cross-sectional surveys. *The ANNALS of the American Academy of Political and Social Science*, *645*, 36–59.

De Leeuw, E., & De Heer, W. (2002). Trends in household survey nonresponse: a longitudinal and international comparison. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds), *Survey nonresponse* (pp. 41–54). New York: Wiley.

Ernst Stähli, M., Joye, D., Pollien, A., Sapin, M. (2012). Enquête non-réponse de MOSAiCH 2011, Centre de compétences suisse en sciences sociales - FORS, Lausanne. Distributed by FORS, Lausanne

Ernst Stähli, M., Joye, D., Pollien, A., Sapin, M.(2012). MOSAiCH: Enquête sur la santé, l'environnement et la politique - 2011 [Dataset]. Centre de compétences suisse en sciences sociales - FORS, Lausanne. Distributed by FORS, Lausanne.

ESS Round 3: European Social Survey Round 3 Data (2006). Data file edition 3.5. Norwegian Social Science Data Services, Norway – Data Archive and distributor of ESS data.

ESS Round 5: European Social Survey Round 5 Data (2010). Data file edition 3.2. Norwegian Social Science Data Services, Norway – Data Archive and distributor of ESS data.

EVS (2010): European Values Study 2008: Switzerland (EVS 2008). GESIS Data Archive, Cologne. ZA4751 Data file version 2.0.0, doi:10.4232/1.10198

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*(5), 646–675.

Groves, R. M., & Heeringa, S. G. (2006). Responsive design for households survey: tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society, Series A*, *169*, 493–457.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias. *Public Opinion Quarterly*, *72*, 1–23.

Hansen, M. H., & Hurwitz, W. N. (1946). The problem of nonresponse in sample surveys. *Journal of the American Statistical Association*, *41*, 517–529.

Joye, D., Schöbi, N., Pollien, A., Kaenel, C. (2008). Enquête non-réponse de l'ESS 2006.

Joye, D., Pollien, A., M. Sapin, M., Ernst Stähli, M. (2011). Enquête non-réponse de l'ESS 2010, Centre de compétences suisse en sciences sociales - FORS, Lausanne. Distributed by FORS, Lausanne.

Kalton, G., & Flores-Cevantes, I. (2003). Weighting methods. *Journal of Official Statistics, 19*(1), 81–97.

Kalton, G., & Maligalig, D. (1991). A comparison of methods of weighting adjustment for nonresponse. In *Proceedings of the 1991 Annual Research Conference* (pp. 401–428). Washington, DC: U.S. Bureau of the Census. Retrieved from http://babel.hathitrust.org/cgi/pt?id=uc1.32106019243457;view=1up;seq=1.

Kreuter, F. (2010). *Improving surveys with paradata: analytic uses of process information*. New York, NY: Wiley.

Kreuter, F., & Olson, K. (2010). Paradata for nonresponse error investigation. In F. Kreuter (Ed.), *Improving surveys with paradata: analytic uses of process information* (pp. 13–43). New York, NY: Wiley

Little, R. J. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review, 54*, 139–157.

Little, R. J., & Vartivarian, S. (2003). On weighting the rates in non-response weights. *Statistics in Medicine, 22*(9), 1589–1599.

Little, R. J., & Vartivarian, S. (2005). Does weighting for nonresponse increase the variance of survey means? *Survey Methodology, 31*, 161–168.

Lorenc, B., Loosveldt, G., Mulry, M. H., & Wrighte, D. (2013). Understanding and improving the external survey environment of official statistics. *Survey Methods: Insights from the Field*. Retrieved June 21, 2014 from http://surveyinsights.org/?p=161.

Lynn, P. (2003). PEDASKI: methodology for collection about survey non-respondents. *Quality & Quantity, 37*, 239–261.

Lynn, P., & Nicolaas, G. (2010). Making good use of survey paradata. *Survey Practice, 3*(2). Retrieved June 21, 2014 from http://surveypractice.org/index.php/SurveyPractice/article/view/127/html.

Matsuo, H., & Billiet, J. (2009). Measurement and correction of nonresponse bias: analysis based on observable information collected from Contacts Forms ESS Round 3. *CeSo: working paper*, CeSO/SM/2009-11.

Matsuo, H., Billiet, J., Loosveldt, G., Berglund, F., & Kleven, O. (2010). Measurement and adjustment of nonresponse bias based on nonresponse survey: the case of Belgium and

Pollien, A., & Joye, D. (2014). Patterns of contact attempts in surveys. In P. Blanchard, F. Bühlmann, & J.-A. Gauthier (Eds.), *Advances in sequence analysis: theory, method, applications* (Chapter 15). London: Springer.

Roberts, C., Lipps, O., & Kissau, K. (2013). Using the Swiss population register for research into survey methodology. *FORS Working Paper Series, paper 2013-1*. Lausanne: FORS. Retrieved from http://forscenter.ch/wp-content/uploads/2013/10/FORS_WPS_2013-01_Roberts-2.pdf.

Roberts, C., Vandenplas, C., & Ernst Staehli, M. (2014). Evaluating the impact of response enhancement methods on the risk of nonresponse bias and survey. *Survey Research Methods*, *8*(2), 67–80.

Saris, W. E., & Gallhofer, I. (2007). *Design, evaluation, and analysis of questionnaires for survey research*. New York: Wiley.

Saris, W. E., Oberski, D., Revilla, M., Zavala, D., Lilleoja, L., Gallhofer, I., & Gruner, T. (2011). The development of the program SQP 2.0 for the prediction of the quality of survey questions. *RECSM Working Paper Number 24*. Retrieved from http://www.upf.edu/survey/_pdf/RECSM_wp024.pdf

Särndal, C.-E., & Lundström, S. (2010). Design for the estimation: identifying auxiliary vectors to reduce nonresponse bias. *Survey Methodology*, *36*, 131–144.

Schneider, S. L., & Kogan, I. (2008). *The International Standard Classification of Education (ISCED-97): an evaluation of content and criterion validity for 15 European countries*. Mannheim: MZES. Retrieved from http://www.mzes.uni-mannheim.de/publications/misc/isced_97/schn08b_the_international_standard_classification_of_educa.pdf.

Schouten, B. (2007). Selection strategy for weighting variables under a not-missing-at-random assumption. *Journal of Official Statistics*, *23*, 51–68.

Schouten, B., Schlomo, N., & Skinner, C. (2011). Indicators for monitoring and improving representativness of response. *Journal of Official Statistics*, *27*, 231–253.

Sinibaldi, J., Trappmann, M., & Kreuter, F. (2014). Which is the better investment for nonresponse adjustment: purchasing commercial auxiliary data or collecting interviewer observations? *Public Opinion Quarterly*, *78*(2), 440–473. doi: 10.1093/poq/nfu003.

Vannieuwenhuyze, J., & Loosveldt, G. (2013). Evaluating relative mode effects in mixed-mode surveys: three methods to disentangle selection and measurement effect. *Sociological Methods and Research*, *42*(1), 82–104.

Wagner, J., & Hubbard, F. (2014). Producing unbiased estimates of propensity models during data collection. *Journal of Survey Statistics and Methodology*, *2*, 323–342.

West, B. T. (2013). An examination of the quality and utility of interviewer's observation in the National Survey of Family Growth. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *176*(1), 211–225.

Acknowledgements

The Swiss European Social Survey, the European Values Study, and the Measurement and Observation of Social Attitudes Survey are financed by the Swiss Science Foundation. We would like to thank our colleagues Michael Ochsner and Marlène Sapin for the fruitful discussions about the subject, their input for this paper, and their work on these surveys. We would also like to thank two anonymous reviewers for their constructive comments that helped to improve the paper.